

Analisis Sentimen Mahasiswa Terhadap Kuliah Offline pada Twitter di Masa Transisi Pandemi COVID-19 Menggunakan Metode Convolutional Neural Network

Mochammad Adhi Buchori¹, Rashif Candra Zirnikh², Febby Milani³

Nur Afifah Az-Zahra⁴, Yaasintha La Jopin Arisca Corpputy⁵

Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta

¹2010511028@mahasiswa.upnvj.ac.id

²2010511031@mahasiswa.upnvj.ac.id

³2010511060@mahasiswa.upnvj.ac.id

⁴2010511085@mahasiswa.upnvj.ac.id

⁵2010511091@mahasiswa.upnvj.ac.id

Abstract— Since March 2020, learning and teaching activities are shifted from conventional learning to online learning. This sudden change to the educational system has caused much controversy. However, after the Covid-19 pandemic has passed, Indonesia is now entering a recovery period. During this transition period, the Government then plans to reopen educational facilities to resume face-to-face learning. However, this decision again caused controversy on social media, including Twitter. This study aims to analyze public opinion on the policy of the return of face-to-face learning during the transition period in Indonesia, especially among students. In this study, the authors conducted text data mining on Twitter social media with the keyword "offline lectures". The dataset obtained was then analyzed and processed using the Convolutional Neural Network (CNN) algorithm. Based on the results of the analysis, implementation and evaluation of the CNN algorithm model, the study resulted in a classification performance with an accuracy value of 76%. The study also concluded that negative sentiment towards offline lectures which are planned to be implemented during the transition period of the Covid-19 pandemic has a higher frequency than positive sentiment.

Keywords— sentiment analysis; Twitter; offline lectures; convolutional neural network.

Abstrak— Sejak Maret 2020, kegiatan belajar dan mengajar dialihkan dari pembelajaran konvensional menjadi pembelajaran daring. Perubahan mendadak terhadap sistem pendidikan tersebut telah banyak menimbulkan kontroversi. Namun, setelah pandemi Covid-19 telah dilalui, kini Indonesia memasuki masa pemulihan. Di masa transisi ini, Pemerintah kemudian berencana membuka kembali sarana pendidikan untuk kembali melakukan pembelajaran tatap muka. Namun, keputusan ini kembali menimbulkan kontroversi di media sosial, termasuk Twitter. Penelitian ini bertujuan untuk menganalisis opini publik terhadap kebijakan kembalinya pembelajaran tatap muka pada masa transisi di Indonesia, khususnya dari kalangan mahasiswa. Pada penelitian ini, penulis melakukan penambangan data teks pada media sosial Twitter dengan kata kunci "kuliah offline". Dataset yang

diperoleh kemudian dianalisis dan diolah menggunakan metode algoritma Convolutional Neural Network (CNN). Berdasarkan hasil analisis, implementasi dan evaluasi model algoritma CNN, penelitian menghasilkan performansi klasifikasi dengan nilai akurasi mencapai 72%. Penelitian juga menyimpulkan bahwa sentimen negatif terhadap perkuliahan offline yang rencananya akan diterapkan di masa transisi pandemi Covid-19 mempunyai frekuensi yang lebih tinggi dibandingkan dengan sentimen positifnya.

Kata kunci— analisis sentimen; Twitter; kuliah offline; convolutional neural network.

I. PENDAHULUAN

Pandemi Covid-19 telah membawa dampak sangat besar terhadap segala aktivitas manusia di dunia. Telah lebih dari dua tahun virus tersebut memasuki Indonesia terhitung sejak Maret 2020. Untuk mengatasi penyebaran virus Covid-19, Pemerintah Indonesia telah banyak mengeluarkan kebijakan dengan mengurangi aktivitas publik sehari-hari salah satunya terhadap sistem pendidikan. Sejak Maret 2020, kegiatan belajar dan mengajar dialihkan dari pembelajaran konvensional menjadi pembelajaran daring. Perubahan mendadak terhadap sistem pendidikan tersebut telah banyak menimbulkan kontroversi. Namun, setelah pandemi Covid-19 telah dilalui, kini Indonesia memasuki masa pemulihan. Di masa transisi ini, Pemerintah kemudian berencana membuka kembali sarana pendidikan untuk kembali melakukan pembelajaran tatap muka. Namun, keputusan ini kembali menimbulkan kontroversi di media sosial, termasuk Twitter. Perubahan mendadak ini kembali menyebabkan berbagai respons dari masyarakat khususnya mahasiswa. Tidak sedikit mahasiswa yang mengeluh mengenai sistem pembelajaran yang dijadikan offline kembali karena beberapa dari mereka yang sudah

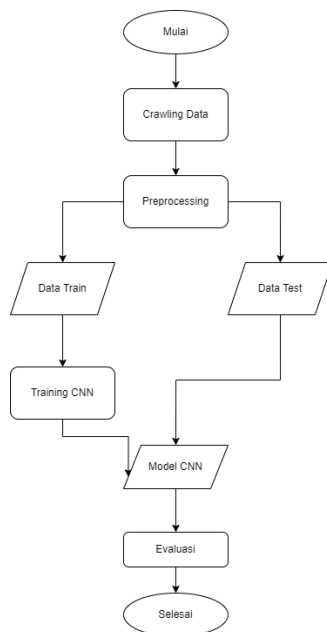
nyaman dengan sistem pembelajaran yang dilakukan secara *online*.

Penelitian ini dibuat untuk menganalisis opini publik terhadap kebijakan kembalinya pembelajaran tatap muka pada masa transisi di Indonesia, khususnya dari kalangan mahasiswa. Pada penelitian ini, penulis menggunakan kata kunci “kuliah offline” yang diambil dari media sosial, Twitter dan diolah menggunakan teknik *text preprocessing* yang kemudian diklasifikasikan menggunakan algoritma CNN untuk mendapatkan analisis sentimen terhadap kebijakan kembalinya pembelajaran tatap muka pada masa transisi di Indonesia.

Penelitian analisis sentimen dengan menggunakan algoritma CNN telah banyak dilakukan sebelumnya. Salah satunya penelitian yang dilakukan Nina I. P. dkk, mengenai analisis sentimen masyarakat terhadap aplikasi e-learning selama pandemi Covid-19 tahun 2021 pada media sosial Twitter dan komentar pada web play store yang menghasilkan performansi klasifikasi dengan nilai akurasi mencapai 86% [4]. Sementara itu, penelitian yang dilakukan oleh Saleh H. B., dkk, mengenai analisis sentimen pelanggan terhadap produk *Indihome* dan *First Media* tahun 2021 pada media sosial *Twitter* yang menghasilkan akurasi tertinggi sebesar 98% untuk provider *IndiHome* dan 91% untuk provider *First Media* [2].

II. METODOLOGI

Proses metodologi penelitian diawali dengan *crawling* data, yaitu proses pencarian *tweet* menggunakan kata kunci yang sudah ditentukan, yakni “kuliah offline”. Alur proses penelitian dapat dilihat di gambar 1.



Gambar. 1 Perancangan Sistem

A. Analisis Sentimen

Analisis Sentimen merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang bertujuan membentuk sistem untuk memahami dan mengekstrak opini dalam bentuk teks. Analisis sentimen atau yang juga dapat disebut sebagai opini *mining* adalah proses penentuan sentimen dalam teks atau kalimat agar dapat dikategorikan sebagai sentimen positif atau negatif [3].

Pada prosesnya, teks atau kalimat yang dianalisis diperoleh dari berbagai sumber data dari internet dan beragam platform media sosial, salah satunya Twitter. Penambangan data opini yang dilakukan bertujuan untuk menganalisis, memproses, dan mengekstrak data tekstual secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini, seperti layanan, produk, individu, hingga fenomena atau topik tertentu [1]. Dengan bantuan analisis sentimen, informasi yang sebelumnya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur.

B. Dataset

Dataset merujuk pada kumpulan data dari informasi lama atau sebelumnya untuk dikelola dengan tujuan menghasilkan informasi baru. Pada umumnya, dataset berisi lebih dari satu variabel dan menyangkut suatu topik tertentu yang digunakan untuk klasifikasi dengan metode data *mining*. Dataset yang digunakan pada penelitian ini merupakan kumpulan *tweet* pengguna Twitter terkait kuliah *offline*. Dataset yang dikumpulkan adalah hasil dari proses *crawling* dengan kata kunci “kuliah offline”.

C. Text Mining

Text mining merupakan proses penambangan data berupa teks untuk mendapatkan informasi yang berguna untuk tujuan tertentu dari berbagai sumber data teks berbeda yang tidak terstruktur [3]. Dengan *text mining*, informasi yang diperlukan dapat diekstrak secara efektif melalui pemrosesan, pengelompokan, dan analisis data-data tidak terstruktur dari sejumlah dokumen dalam jumlah besar.

D. Text Preprocessing

Text Preprocessing pada penelitian ini merupakan tahapan untuk mengubah struktur data yang didalamnya terdapat kumpulan *text* menjadi data *clean text*. *Preprocessing* dilakukan agar saat dilakukan klasifikasi, hasil proses uji coba dapat menampilkan hasil akurasi yang tinggi. Data *clean text* merupakan data yang sudah seragam dan tidak mengandung data yang tidak relevan.

- Cleaning* Data merupakan tahapan menghapus *mention* (@), *hashtag*, *retweet*, URL, angka, mengganti baris baru menjadi spasi, dan membuang tanda baca.
- Case folding* merupakan tahapan mengubah semua huruf dalam suatu kalimat menjadi huruf kecil. Setiap kalimat tidak konsisten dalam penggunaan huruf

kecil dan kapital. Oleh karena itu, pengubahan huruf kapital menjadi huruf kecil diperlukan.

- c. *Tokenizing* merupakan proses mengubah sebuah dokumen menjadi bagian-bagian, seperti sebuah token. Dalam hal ini, setiap kalimat dipisahkan menjadi kata satu per satu.
- d. *Filtering* kata atau melakukan *stopword removal* merupakan tahapan untuk menghilangkan kata-kata yang tidak diperlukan, seperti konjungsi atau kata yang tidak mengandung makna.
- e. *Stemming* merupakan tahap dimana kata-kata diubah ke dalam bentuk kata dasar dengan menghapus imbuhan awal, akhir, atau keduanya.

E. Labelling Data

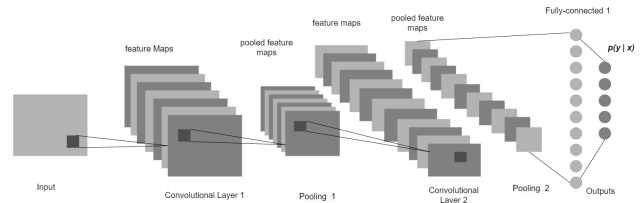
Labelling data merupakan proses penandaan pada suatu data utuh, seperti gambar, video, teks ataupun audio. Proses *labelling* data pada penelitian ini menggunakan data utuh teks yang dikumpulkan dari *Twitter*. Penandaan ini membentuk representasi nilai sentimen dari data *tweet* yang telah dikumpulkan. *Labelling* data pada penelitian ini dilakukan secara manual oleh dua orang peneliti agar mendapatkan hasil dua perspektif.

F. Word Cloud

Word Cloud merupakan sebuah sistem yang dapat menggambarkan ataupun memvisualisasikan kata-kata dengan memperlihatkan frekuensi kata yang digunakan dalam suatu teks tertulis. Sistem tersebut digunakan dengan tujuan untuk memperoleh informasi tentang opini publik terhadap kebijakan kembalinya pembelajaran tatap muka pada masa transisi pandemi Covid-19 di Indonesia. Data yang diaplikasikan pada *Word Cloud* adalah data hasil *labelling* dan *preprocessing*.

G. Metode Convolutional Neural Network

Convolutional Neural Network atau CNN merupakan metode untuk memproses fakta yang mempunyai topologi jaringan. CNN mengindikasikan bahwa jaringan tersebut menggunakan operasi matematika konvolusi. CNN merupakan salah satu algoritma *deep learning* yang umumnya terdiri dari 3 lapisan utama, yaitu *input layer*, *output layer*, dan beberapa lapisan penyusun yang saling terhubung untuk memproses data *input*. Lapisan penyusun tersebut terdiri dari *convolution layer* dan *pooling layer* sebagai lapisan ekstraksi fitur yang digunakan untuk memproses dan mempelajari fitur pada data *input*, serta lapisan *fully connected layer* yang digunakan untuk melakukan proses klasifikasi pada data *input* yang telah diekstrak fiturnya.



Gambar. 2 Ilustrasi *Convolutional Neural Network*

H. Evaluasi

Model yang telah dibangun akan dilakukan evaluasi dengan menggunakan metode *Confusion Matrix*. Evaluasi dengan menggunakan *confusion matrix* ini akan menghasilkan nilai dari akurasi, presisi, recall, dan kappa untuk mengukur kesuksesan dari model algoritma *Convolutional Neural Network* (CNN).

- a. *Confusion matrix* merupakan metode validasi data teks dari jumlah opini negatif dan positif yang tidak seimbang.

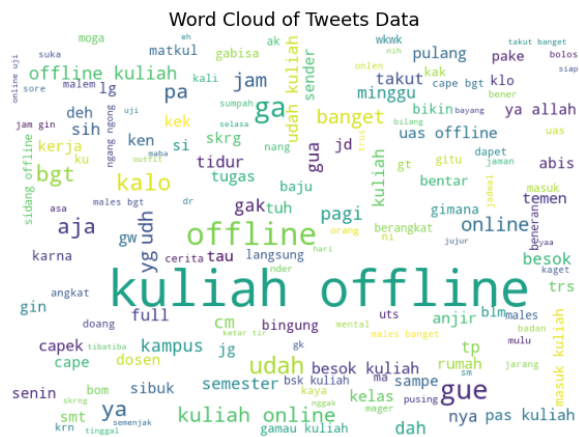
TABEL I
CONFUSION MATRIX

Confusion Matrix		Predicted Value	
		Positive	Negative
Actual Value	Positive	TP	FN
	Negative	FP	TN

- i. Nilai Prediksi adalah keluaran dari program dimana nilainya Positif dan Negatif.
- ii. Nilai Aktual adalah nilai sebenarnya dimana nilainya *True* dan *False*.
- iii. True Positive (TP) adalah keluaran program dimana nilainya Positif dan bersifat *True*.
- iv. True Negative (TN) adalah keluaran program dimana nilainya Negatif dan bersifat *True*.
- v. False Positive (FP) adalah keluaran program dimana nilainya positif dan bersifat *False*.
- vi. False Negative (FN) adalah keluaran program dimana nilainya Negatif dan bersifat *False*.
- b. Akurasi merupakan rasio antara prediksi yang benar dengan seluruh sampel yang diprediksi.

$$Akurasi = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \times 100\% \quad (1)$$

- c. *Recall* merupakan rasio antara prediksi positif yang benar dengan seluruh sampel positif (*Actual*).



Gambar. 7 Visualisasi Word Cloud Negatif

F. Klasifikasi

Algoritma yang digunakan dalam proses klasifikasi, yakni algoritma *Convolutional Neural Network* (CNN). Dalam penerapannya, klasifikasi sentimen dilakukan dengan dataset hasil *preprocessing* dengan jumlah total sebanyak 1637 data. Data tersebut kemudian dibagi menjadi data *train* dan data *test* dengan pembagian sebanyak 80% atau 1309 data untuk data *train* dan 20% atau 328 data untuk data *test*.

Sebelum membangun model CNN, penulis melakukan proses vektorisasi berupa *one hot encoding* menjadi nilai numerik dimana sentimen positif setara dengan 0 dan sentimen negatif setara dengan 1. Setelahnya, penulis melakukan pemodelan menggunakan algoritma *Convolutional Neural Network* (CNN) menggunakan model Sequential. Dalam implementasinya, penulis menggunakan 4 jenis layer, yakni *Embedding*, *LSTM*, *Dense*, dan *Dropout*. Dalam hal ini, model tersebut penulis implementasikan dalam *wrapper* KerasClassifier. Adapun isi parameter dari model tersebut yang dapat dilihat pada tabel berikut:

TABEL IX
LAYER DAN PARAMETER MODEL SEQUENTIAL

Layers	Parameter	Value
Embedding	input_dim	5000
	output_dim	32
	input_length	24
LSTM	units	16
	activation	tanh
Dropout	rate	0.2
Dense	units	3
	activation	softmax

Layer pertama, yakni *Embedding*. Layer tersebut berfungsi untuk mengubah bilangan bulat positif (indeks) menjadi vektor dengan ukuran tetap. Layer selanjutnya, yaitu *Long Short-Term Memory* (LSTM) yang berfungsi untuk mengklasifikasikan teks berdasarkan pada urutan kata dalam sebuah kalimat. Layer ketiga, yakni *Dropout*

yang berfungsi untuk mencegah terjadinya *overfitting*. Layer terakhir, yakni *Dense* yang berfungsi sebagai hidden layer dan output layer pada sebuah Multi Layer Perceptron (MLP).

Model CNN dibuat menggunakan *loss function* guna menemukan kesalahan dalam proses pembelajaran. *Loss function* yang digunakan, yaitu *sparse_categorical_crossentropy* karena model yang dibangun merupakan klasifikasi antara 2 kategori dengan menggunakan tipe data integer dalam menyimpan label. Lalu, penulis juga menggunakan *optimizer* guna mengoptimalkan bobot input dengan membandingkan prediksi dengan *loss function*. *Optimizer* yang digunakan, yaitu Adam dimana merupakan *optimizer* dengan akurasi terbaik yang dipilih berdasarkan percobaan terhadap beberapa *optimizer* lainnya. Dalam hal ini, *optimizer* tersebut juga dikombinasikan dengan *learning rate* sebesar 0.01. Terakhir, penulis menggunakan *metrics* guna mengevaluasi performa model. Dalam hal ini, *metrics* yang digunakan, yaitu *accuracy*. Informasi lengkap dari pemilihan parameter tersebut dapat dilihat pada tabel berikut:

TABEL X
PARAMETER COMPILE MODEL

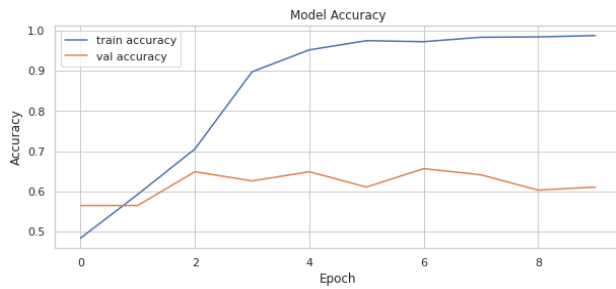
Layers	Value
loss	sparse categorical crossentropy
optimizer	Adam
learning rate	0.01
metrics	accuracy

Di samping itu, dalam pembuatan model, penulis juga menggunakan fit parameters berupa epochs sebanyak 10, batch size sebesar 128 dan validation split sebesar 10% atau 0.1. Adapun kesimpulan dari model yang dibangun yang dapat dilihat pada tabel berikut:

TABEL XI
KESIMPULAN MODEL

Layers	Output Shape	Param
Embedding	None, 24, 32	160000
LSTM	None, 16	3136
Dropout	None, 16	0
Dense	None, 3	51

Berikut merupakan performa dari model CNN yang dibangun yang penulis representasikan dalam bentuk grafik:



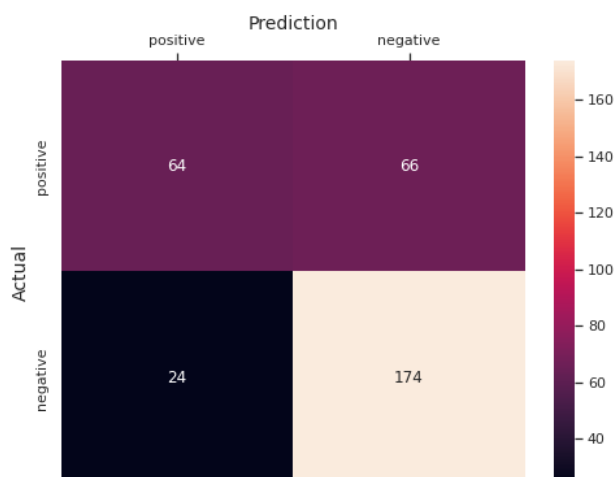
Gambar. 8 Grafik Plot Accuracy

TABEL IX
ITERASI EPOCH

Epoch	Nilai Train Accuracy	Nilai Validation Accuracy
1	0.4839	0.5649
2	0.5925	0.5649
3	0.7054	0.6489
4	0.8973	0.6260
5	0.9516	0.6489
6	0.9745	0.6107
7	0.9720	0.6565
8	0.9830	0.6412
9	0.9839	0.6031
10	0.9873	0.6107

G. Evaluasi

Setelah tahap klasifikasi selesai, selanjutnya penulis mengevaluasi model terhadap algoritma *Convolutional Neural Network* (CNN) yang dibentuk. Evaluasi tersebut dilakukan dengan menghitung nilai akurasi, *recall*, *precision*, dan kappa. Perhitungan tersebut dipacu berdasarkan *confusion matrix* yang dibentuk dengan nilai aktual berupa data test dan nilai prediksi hasil pemodelan. Berikut merupakan ilustrasi dari *confusion matrix* tersebut:



Gambar. 10 Confusion Matrix

Berdasarkan gambar di atas, dapat diketahui bahwa *confusion matrix* dari hasil pemodelan terdiri dari 64 *true positive*, 174 *true negative*, 24 *false positive*, dan 66 *false negative*. Dengan demikian, model tersebut mempunyai nilai akurasi sebesar 0.7256097561, nilai *recall* sebesar 0.7272727273, nilai presisi sebesar 0.4923076923, dan nilai kappa sebesar 0.4815595364.

IV. KESIMPULAN

Penelitian ini menerapkan algoritma *Convolutional Neural Network* untuk klasifikasi sentimen. Analisis sentimen dalam studi ini menggunakan data Twitter berupa *tweet* dengan kata kunci “kuliah offline” dalam bahasa Indonesia yang diambil mulai dari tanggal 22 Mei 2022 hingga 29 Mei 2022. Penilaian ini menghasilkan nilai akurasi sebesar 0.7256097561, nilai *recall* sebesar 0.7272727273, nilai presisi sebesar 0.4923076923, dan nilai kappa sebesar 0.4815595364. Dalam hal ini, beberapa *tweet* menunjukkan ketidaksiapan untuk kuliah *offline* dengan representasi kata ‘capek’ dan ‘bingung’ yang memiliki frekuensi tinggi dalam *tweet*. Dengan demikian, berdasarkan hasil penelitian, dapat disimpulkan bahwa sentimen negatif terhadap perkuliahan *offline* yang rencananya akan diterapkan di masa transisi pandemi Covid-19 mempunyai frekuensi yang lebih tinggi dibandingkan dengan sentimen positifnya.

REFERENSI

- [1] B. Liu. 2012. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool. Publishers.
- [2] Badjrie, S. H., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis Sentimen Review Customer Terhadap Produk Indihome Dan First Media Menggunakan Algoritma Convolutional Neural Network Review Analysis Sentiment Customer Product Indihome And First Media Using Convolutional Neural Network. 8(5), 9049–9061.
- [3] Samsir, Ambiyar, Unung, V., Firman, E., & Ronal, W. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. Jurnal Media Informatika Budidarma, 5(1), 157-163. <https://doi.org/10.30865/mib.v5i1.2604>
- [4] Simbolon, A. S., Pangaribuan, N. I., & Aruan, N. M. (2021). Analisis Sentimen Aplikasi E-Learning Selama Pandemi Covid-19 Dengan Menggunakan Metode Support Vector Machine Dan Convolutional Neural Network. Seminastika, 3(1), 16–25. <https://doi.org/10.47002/seminastika.v3i1.236>

KONTRIBUSI ANGGOTA KELOMPOK

No	Nama	NIM	Kontribusi
1	Mochammad Adhi Buchori	2010511028	<ol style="list-style-type: none">1. Membantu proses pengumpulan dataset.2. Membuat, melakukan riset, dan mencari referensi kode program untuk penelitian.3. Menyusun laporan atau jurnal dengan format IEEE.4. Membuat slide presentasi untuk dipresentasikan.5. Mengumpulkan referensi yang mendukung penyusunan laporan atau jurnal penelitian.
2	Rashif Candra Zirnikh	2010511031	<ol style="list-style-type: none">1. Membantu proses pengumpulan dataset.2. Membuat, melakukan riset, dan mencari referensi kode program untuk penelitian.3. Menyusun laporan atau jurnal dengan format IEEE.4. Membuat slide presentasi untuk dipresentasikan.5. Mengumpulkan referensi yang mendukung penyusunan laporan atau jurnal penelitian.
3	Febby Milani	2010511060	<ol style="list-style-type: none">1. Membantu proses pengumpulan dataset.2. Membuat, melakukan riset, dan mencari referensi kode program untuk penelitian.3. Menyusun laporan atau jurnal dengan format IEEE.

			<ol style="list-style-type: none"> 4. Membuat slide presentasi untuk dipresentasikan. 5. Mengumpulkan referensi yang mendukung penyusunan laporan atau jurnal penelitian.
4	Nur Afiifah Az-Zahra	2010511085	<ol style="list-style-type: none"> 1. Membantu proses pengumpulan dataset. 2. Membuat, melakukan riset, dan mencari referensi kode program untuk penelitian. 3. Menyusun laporan atau jurnal dengan format IEEE. 4. Membuat slide presentasi untuk dipresentasikan. 5. Mengumpulkan referensi yang mendukung penyusunan laporan atau jurnal penelitian.
5	Yaasintha La Jopin Arisca Corpputy	2010511091	<ol style="list-style-type: none"> 1. Membantu proses pengumpulan dataset. 2. Membuat, melakukan riset, dan mencari referensi kode program untuk penelitian. 3. Menyusun laporan atau jurnal dengan format IEEE. 4. Membuat slide presentasi untuk dipresentasikan. 5. Mengumpulkan referensi yang mendukung penyusunan laporan atau jurnal penelitian.

DOKUMENTASI

Remaining Meeting Time: 04:17 | Upgrade to Pro

View

PROJECT AKHIR
TEMU KEMBALI INFORMASI

2010511031_Rashif Candra Zirnikh

PROJECT AKHIR
TEMU KEMBALI INF

2010511028_Mochammad Adhi Buchori

PROJECT AKHIR
TEMU KEMBALI INFORMASI

Febby Milani

PROJECT AKHIR
TEMU KEMBALI INFORMASI

2010511085_Nur Afifah Az-Zahra

PROJECT AKHIR
TEMU KEMBALI INFORMASI

Yaasintha La Jopin Arisca Corpputy

Yaasintha La Jopin Arisca Corpputy

Mute Stop Video Security Participants 6 Chat Share Screen Record Reactions Apps Whiteboards End